

METHOD AND APPARATUS FOR NETWORK LOAD-BALANCINGTECHNICAL FIELD

The invention relates to the field of communications systems and, more specifically, to an adaptive/iterative load-balancing method suitable for use in network management systems providing automatic route provisioning and/or manual route provisioning.

BACKGROUND OF THE INVENTION

Telecommunication networks and other networks are increasing in both size and complexity. It is anticipated that this trend will continue such that very large telecommunications networks having tens of thousands of nodes will become increasingly commonplace. Unfortunately, as such networks increase in size, the network management function also increases in complexity. This means that critical tasks such as provisioning (allocating resources to form a communications link), restoration, reinstatement and the like, must be completed in a reasonable time using network management tools available to a network manager at a single location.

In a manual provisioning mode, an operator specifies all details of a circuit such as end points, all links, time slots, and all network elements. The manual provisioning mode allows the operator to select a particular circuit providing a communication circuit for DS-1, DS-3, EC-1, OC-3 and other communications services. However, the manual provisioning mode is slow (the operator must select all links manually) and error prone (the operator may make an error in selecting these links).

In an automatic provisioning mode, the operator specifies end points (i.e., start node and end node) and type of circuit needed to provide the desired communication. A network manager system responsively examines all of the spare resources available in the network and selects the optimum path for the

5

10

20

SUMMARY OF THE INVENTION

25

30

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts a high level block diagram of a communications system;

5 FIG. 2 depicts a high level block diagram of a network manager suitable for use in the communications system of FIG. 1;

FIG. 3 depicts a graphical representation of a source node, an end node and a plurality of intervening nodes within a
10 network;

FIG. 4 comprises a flow diagram of a method of selecting a load-balanced shortest path according to an embodiment of the invention.

To facilitate understanding, identical reference numerals
15 have been used, where possible, to designate identical elements that are common to the figures.

DETAILED DESCRIPTION OF THE INVENTION

The subject invention will be described within the context
20 of a telecommunication system comprising a large number of network elements or nodes interconnected in a mesh topology. However, it will be appreciated by those skilled in the art that the subject invention may be advantageously employed in any communications network in which provisioning of any form of
25 communication may be utilized, such as telecommunication, data communication, streaming media communication and the like. Thus, it is contemplated by the inventors that the subject invention has broad applicability beyond the telecommunication network described herein.

30 Provisioning (manual or automatic) comprises the process of selecting the start and end points (nodes) of a communication path, selecting all the nodes and links connecting the start and end nodes, finding the "best" communication path between the start and end nodes, and
35 generating the commands to each of the nodes within the "best"

path such that cross-connects within the network cause the path to be formed, thereby enabling traffic flow through the provisioned circuit. It is noted that each link typically comprises a plurality of channels, and that each channel typically has time slots that can be reserved.

The provisioning of a circuit is a network management layer function within the telecommunications management network (TMN) standards, described in more detail in International Telecommunications Union (ITU) standard documents such as recommendation M.3010 and related documents, which are incorporated herein by reference in their entireties. It is noted that the TMN functional layers also include a service management layer which is above, and interacts with, the network management layer. Therefore, in the case of TMN management at the service management layer, the network management layer functions may not be performed manually.

FIG. 1 depicts a high level block diagram of a communications system. Specifically, the communications system 100 of FIG. 1 comprises a database 110, a network manager or controller 120, a plurality of work stations 130₁ through 130_n (collectively work stations 130), a communications link 134 and a multi-node communication network 140.

The multi-node communication network 140 comprises a plurality of network elements (NE) denoted as network elements NE₁ through NE_x (collectively network elements NE). Also depicted is a start-node SN and an end-node EN. As will be discussed in more detail below with respect to FIGS. 3 and 4, the invention operates to determine the shortest path between a source (start) node and a destination (end) node for provisioning a circuit in a manner that adapts to the bandwidth utilization or "loading" placed upon each link connecting the source node, destination node, and intervening nodes. In this manner, the invention operates to insure that each link within the provisioned circuit is operating at a loading level below a threshold value. It is noted that the threshold level may be

adapted (i.e., increased or decreased) in response to, for example, calculated circuit paths being too long (e.g., a calculated circuit path comprising 50% more nodes than a circuit path calculated without respect to link loading conditions).

The multi-node communication network 140 is coupled to the network manager 120 via signal path S3. The network manager or controller 120 is used to manage various network operations such as the routing of communications and other functions. Specifically, in one embodiment, the multi-node communication network 140 comprises a large number of network elements where each communication to be transmitted from a start network element or start-node to an end network element or end-node requires the determination by the network manager 120 of an appropriate communications path.

The database 110 may comprise a standard mass storage device, such as a redundant array of inexpensive devices (RAID) or other known mass storage device cooperating with a data base program such as the Oracle data base provided by Oracle Corporation of Redwood Shores, California. All that is necessary is that the database 110 be able to communicate with the network manager 120 in a manner facilitating the storage and retrieval of information, such as characterization and control information pertaining to the multi-node communication network 140 including loading information regarding the various links interconnecting the nodes in the network. In one embodiment of the invention, the data base 110 stores information pertaining to each node within the multi-node communication network 140 and, more particularly, to the type of links connecting the nodes, the type of channels provided by these links and the loading or bandwidth utilization of the respective links and/or channels. The data base 110 also stores information pertaining to the availability of time slots for the various links and/or channels used to communicate between nodes.



Each of the work stations 130 may comprise, for example, a terminal used by a network operator to request the provisioning of communication circuits between start-nodes and end-nodes in response to, for example, requests for such circuits from network users. The work stations 130 may also comprise interfaces between network system users and customers and the network manager 120. Within the context of the present invention, the work stations 130 are used to provide information to the network manager or controller 120 indicative of at least the start node and end node of a circuit to be provisioned, as well as any quality of service (QOS) or other transmission parameters associated with that circuit. Broadly speaking, all that is necessary to practice the present invention is a communication from some entity, such as a work station 130, indicative of the start node and end node of a circuit to be provisioned.

The network manager 120 and database 110 of the communications system 100 of FIG. 1 are depicted as separate functional entities. However, it will be appreciated by those skilled in the art that the network manager 120 and database 110 may be combined within a single functional entity. Thus, the network manager 120 and database 110 may be operably combined to form a network management apparatus suitable for managing the multi-node communication network 140 according to the present invention.

30 In one embodiment, the network manager 120 comprises, illustratively, an Integrated Transport Management Network Manager (ITM-NM) manufactured by Lucent Technologies, Inc. of Murray Hill, New Jersey. In this embodiment, the network manager 120 implements network management layer functions
35 according to, for example, the Telecommunications Management

Network (TMN) standards described in the International Telecommunications Union (ITU) recommendation M.3010 and related documents, which are incorporated herein by reference in their entirety. Thus, the network manager 120 is used to
5 manage all network elements within the communications system 100 of FIG. 1, both individually and as a set of network elements. The network manager 120 can include or be operatively coupled to various element management systems EMS_1 through EMS_n (collectively the element management systems EMS)
10 according to the various management layer functions described in the TMN standard.

FIG. 2¹ depicts a high level block diagram of a network manager or controller suitable for use in the communications system 100 of FIG. 1. Specifically, the exemplary network
15 manager or controller 120 of FIG. 2 comprises a processor 120-4 as well as memory 120-8 for storing various network management and control programs 120-8P. The processor 120-4 cooperates with conventional support circuitry 120-3 such as power
20 supplies, clock circuits, cache memory and the like as well as circuits that assist in executing the software routines stored in the memory 120-8. As such, it is contemplated that some of the process steps discussed herein as software processes may be implemented within hardware, for example, as circuitry that
25 cooperates with the processor 120-4 to perform various steps. The network manager 120 also contains input-output (I/O) circuitry 120-2 that forms an interface between the various functional elements communicating with the network manager 120. For example, in the embodiment of FIG. 1, the network manager 120 communicates with a database 110 via a signal path S1, each
30 of a plurality of work stations 130 via signal path S2 and the communication network to be managed 140 via signal path S3.

Although the network manager 120 of FIG. 2 is depicted as a general purpose computer that is programmed to perform various network management functions in accordance with the
35 present invention, the invention can be implemented in hardware

as, for example, an application specific integrated circuit (ASIC). As such, the process steps described herein are intended to be broadly interpreted as being equivalently performed by software, hardware, or a combination thereof.

5 The network manager 120 of the present invention communicates with the various work stations 130 such as those being used by network operators servicing customers requesting specific connections.

FIG. 3 depicts a graphical representation of circuit
10 comprising a source node, an end node and a plurality of intervening nodes within a network. Specifically, FIG. 3 depicts a network 300 comprising a plurality of nodes denoted as nodes A through R. Circuit paths, or links exist between some of the nodes, but not all of the nodes (though each node
15 is coupled to at least two other nodes). Circuit links are denoted by two letters indicative of the nodes connected by the respective circuit link.

In the network 300 of FIG. 3, a circuit is to be provisioned between a start node (node A) and an end node (node
20 G). The shortest path between nodes A and G is the path through the following nodes A-H-I-J-G. However, according to the invention, if the bandwidth utilization level or "load" on any of the links connecting nodes A through G has reached or exceeded a preset threshold level, such as 50% bandwidth
25 utilization level, then another shortest or "next shortest" path is found where the load is found, such that the threshold level is honored. As depicted in FIG. 3, the next shortest path between nodes A and G is the path through nodes A-B-C-D-E-G. It is noted that the first shortest path traverses three
30 intervening nodes (H, I and J), while the next shortest path traverses four intervening nodes (B, C, D and E). The next shortest path A-B-C-D-E-G is formed by coupling nodes A and B via path AB, nodes B and C via path BC, nodes C and D via path CD, nodes D and E via path DE, and nodes E and G via path EG.

With respect to the load balancing aspects of the invention, the threshold level for determining whether a link between the two nodes is over utilized may be predetermined or user settable. Moreover, the threshold level preferably is defined with respect to the type of link joining the two nodes. Additionally, the threshold level preferably applies to each of the digital links used to connect the start node, end node and intervening nodes.

It is noted that the threshold level is a measure for a digital link, not a measure of the aggregate of links between two nodes. That is, the threshold level is applied to the specific digital link between two nodes contemplated to be used within the provisioned circuit. Where multiple links between two nodes exist, alternate links may be used or the multiple links may have associated with them different threshold levels, depending on the technology used to provide each link. In this manner, the "shortest path" algorithm and threshold level comparison are used in an iterative fashion whereby each link determined to be appropriate according to the shortest path algorithm is compared to a corresponding threshold level to determine if the link is, in fact, appropriate with respect to the bandwidth utilization level of the link. If the link is over utilized or otherwise inappropriate, then a different link may be selected for use in the shortest path algorithm. In this manner, those links following an inappropriate or over utilized link do not have to be processed by the shortest path algorithm.

FIG. 4 comprises a flow diagram of a method of selecting a load-balanced shortest path according to an embodiment of the invention. Specifically, the method 400 of FIG. 4 utilizes existing shortest path algorithms in an iterative fashion such that each link of a proposed shortest path is checked for loading prior to inclusion within a provision circuit.

The method 400 of FIG. 4 is entered at step 402 and proceeds to step 404, where all links for a shortest path

determination are accepted. That is, at step 404, none of the possible or acceptable links used to provide a circuit path between a source node and a destination node are excluded from consideration by the shortest path algorithm. The method 400
5 then proceeds to step 406.

At step 406, at least one link of the shortest path between the source node and the destination node is determined using the presently accepted links. In one embodiment of the invention, only a single link (i.e., a next link) extending
10 from the node connected to a previously processed link is determined. In another embodiment of the invention, a larger portion or an entirety of a "shortest path" between the node connected to the processed link and the end node is determined. In the case of step 406 being executed for the first time to
15 determine an entire path, the determined shortest path comprises a "ideal" shortest path. This ideal shortest path comprises the shortest path between the start node and the end node based primarily on the topology of the network and excluding any consideration of the bandwidth utilization levels
20 of the links used to provide such path. The method 400 then proceeds to step 408.

During the process of constructing an acceptable shortest path between the source node and the destination node, a presently calculated "acceptable" path is formed beginning with
25 the source node and proceeding toward the destination node. During the formation of this path, the last or terminal node of a path so formed comprises the last node of a path connected to the source node via one or more accepted links. It is noted that for each iteration of the method, a single next link
30 (coupling the terminal node to a next node) and a plurality of next links (coupling the terminal node to a respective plurality of next nodes or an entire group of links necessary to couple the terminal node to the destination node via as many intervening nodes and links as necessary) may be provided.

At step 408, the loading of the first or next link in the calculated shortest path is determined. That is, if step 408 is being executed for the first time, then the loading of the first link in the determined shortest path is determined. The
5 first link comprises a link between the source node and the first node within the determined shortest path. If step 408 has been previously executed, then the next link loading is determined. For example, in the case of step 408 being executed for the second time, the next link comprises the link
10 bridging the node connected to the first link and the next node. The method 400 then proceeds to step 410.

At step 410, a query is made as to whether the loading of the link determined at step 408 is less than or equal to a threshold level, such as 30%, 50%, 70%, 100% or some other
15 value. If the query at step 410 is answered affirmatively, then the method 400 proceeds to step 414. If the query at step 410 is answered negatively, then the method 400 proceeds to step 412.

At step 412, the link having a loading determined at step
20 408 is rejected for consideration for the circuit being provisioned. The method 400 then proceeds to step 406, where the shortest path between the source node and destination node is determined using accepted links. The second and subsequent executions of step 406 may utilize the links already determined
25 to have loading levels below their respective threshold levels or by recalculating the entire circuit path. Preferably, the method of the present invention is performed by recalculating the shortest path using links that are known to be loaded below their respective threshold levels.

At step 414 the link having a loading level determined at
30 step 408 is accepted for use in the circuit being provisioned. The method 400 then proceeds to step 416, where a query is made as to whether the path is now complete. That is, at step 416, a query is made as to whether the link accepted at step 414
35 comprises the final link between a penultimate node and the

005280" / 0205950

destination or end node. If the query at step 416 is answered negatively, then the method 400 proceeds to step 408. If the query at step 416 is answered affirmatively, then the method 400 proceeds to step 418.

- 5 At step 418, a query is made as to whether the path of the circuit defined by the iterative process described above with respect to FIGS. 406 through 416 is too long. For example, if the path completed using the iterative process is greater than 40%, 50%, 60%, 100% or some other percentage (or distance, latency or other metric) longer than an initially determined or "ideal" shortest path (i.e., a shortest path between the source node and destination node determined without regard to loading), then the path may be deemed to be too long. If the query at step 418 is answered affirmatively, then the method 15 400 proceeds to step 420. If the query at step 418 is answered negatively, then the method 400 proceeds to step 422.

- At step 420, the threshold levels of one or more of the links are adjusted. That is, assuming that a default threshold level of 50% loading has been used to accept or exclude links 20 for purposes of provisioning a circuit and that the resulting circuit path has been deemed to be too long at step 418. In this instance, the threshold level applied to one or more of the links may be increased (or decreased) to any level up to 100%. The actual increase in threshold level is preferably 25 made by examining the type of links available for use in provisioning a circuit and modifying the threshold levels accordingly. It should be noted that the threshold level for each of the links between intervening nodes need not be the same, and that individual threshold levels, groups of threshold 30 levels, or the entirety of the threshold levels associated with the links interconnecting the intervening nodes may be adjusted. The method 400 then proceeds to step 404. If the query at step 418 is answered negatively, then the method 400 proceeds to step 422 where the circuit is provisioned and 35 tested. The method exits at step 424.

Although various embodiments which incorporate the teachings of the present invention have been shown and described in detail herein, those skilled in the art can readily devise many other varied embodiments that still incorporate these teachings.

30 incorporate these teachings.